Spatially Selective Active Noise Control Systems

Tong Xiao,[1, a] Buye Xu,[2, b] and Chuming Zhao[2, c]

[1] *Centre for Audio, Acoustics and Vibration, University of Technology Sydney, NSW 2007, Australia*

[2] *Meta Reality Labs Research, Redmond, WA 98052, USA*

(Dated: 17 April 2023)

Active noise control (ANC) systems are commonly designed to achieve maximal sound reduction regardless of the incident direction of the sound. When desired sound is present, the state-of-the-art methods add a separate system to reconstruct it. This can result in distortion and latency. In this work, we propose a multi-channel ANC system that only reduces sound from undesired directions, and the system truly preserves the desired sound instead of reproducing it. The proposed algorithm imposes a spatial constraint on the hybrid ANC cost function to achieve spatial selectivity. Based on a six-channel microphone array on a pair of augmented eyeglasses, results show that the system minimized only noise coming from undesired directions. The control performance could be maintained even when the array was heavily perturbed. The proposed algorithm was also compared with the existing methods in the literature. Not only did the proposed system provide better noise reduction, but it also required much less effort. The binaural localization cues were not needed to be reconstructed since the system preserved the physical sound wave from the desired source.

[a]Tong.Xiao@uts.edu.au

[b]xub@meta.com

[c]chumingz@meta.com

I. INTRODUCTION

Active noise control (ANC) systems have seen many significant advancements over the past few decades. Notably, many personal ANC headphones, aiming to eliminate the unwanted noise for users, have emerged and gained much success in the market due to their excellent performance and robustness (Chang *et al.*, 2016). Other ANC systems, such as ANC headrests (Elliott *et al.*, 2018; Xiao *et al.*, 2020) and ANC windows (Lam *et al.*, 2020; Wang *et al.*, 2017), have also seen significant improvements over the years. These ANC systems have been designed to attenuate all the sound in the environment. One emerging application is using ANC to enhance face-to-face conversations in a noisy environment, e.g., a cocktail-party scenario, where one hopes to minimize the surrounding noises while still maintaining the conversations in front of the user. An intelligent ANC system should separate and categorize sounds coming from various directions. The desired sound should be maintained for the user while noises coming from other directions are minimized.

The goal of this work is to develop a spatially selective ANC system that preserves sound coming from the desired directions while reducing noise from other directions. The filtered-reference least-mean-square (FxLMS) algorithm is commonly used in these systems for adaptive control due to its simplicity and robustness (Elliott, 2000; Hansen *et al.*, 2012; Kuo and Morgan, 1996). Any signals observed by the reference microphones (or error microphones in the feedback systems) will be fully controlled regardless of the residual noise spectrum at the error microphones. So far, most effort in the ANC systems has been devoted to improving the noise reduction level throughout the spectrum as much as possible.

In recent years, some studies have considered the spatial aspect of personal ANC systems. Studies (Cheer *et al.*, 2019; Liebich *et al.*, 2018; Rafaely and Jones, 2002; Zhang and Qiu, 2014) have showed that the performance in ANC headphones and earphones was affected by the time advance between the reference and the error microphones in the feedforward configurations. The ANC performance was the best when the time advance was the most significant. However, these systems were studied to further improve the control performance when dealing with direction- and/or time-varying noises. They were not designed for spatial noise selection.

There are systems, which were not intentionally designed, that can be modified to achieve spatial selection functionality. For example, the coherence-based selection method (Shen *et al.*, 2021) can potentially be used to determine the direction of arrival (DOA) of the noise, which can then be isolated. The selective fixed-filter method (Shi *et al.*, 2020, 2022) and the deep ANC method (Zhang and Wang, 2021) can be extended to selecting spatial filters to control noise coming from certain DOAs. However, these systems still possess issues. The coherence-based selection method requires the input signals to be distinguished enough to offer differences in the coherence functions. Thus, it may be limited to certain spread-out array configurations, e.g., an array distributed in a room. The selective fixed-filter method and the deep ANC method require the system to be pre-trained to obtain the optimal filters in advance. These two methods also require the spatial information to be acquired elsewhere and thus resulting in additional computations. These systems are not yet the best solutions for a spatially selective ANC system.

To achieve spatial selectivity, it is common to use the beamforming technique. Beamforming is a well-established method of designing spatio-temporal filters for array processing (Doclo *et al.*, 2015; Van Trees, 2002; Van Veen and Buckley, 1988). For personal devices, a certain number of microphones can be used to differentiate sound from different directions. Hearing aids typically employ it to improve speech intelligibility for the hearing impaired. Studies have incorporated the ANC functionality to control the leakage sound in open-fitting systems, which improves the insertion signal-to-noise-ratio (SNR) gain of the multi-channel Wiener filter (MWF) (Dalga and Doclo, 2011; Serizel *et al.*, 2010). Serizel *et al.* (2010) used the feedforward ANC to control the leakage noise, while Dalga and Doclo (2011) further improved the result by using the hybrid control due to its better ANC performance. However, the desired sound in the leakage is canceled (together with the noise) and then added back to the error signal with a delay. Recent work in (Patel *et al.*, 2020) proposed a structure for a pair of hear-through ANC headphones. The microphones were grouped for different purposes independently, i.e., ANC and beamforming, and the desired sound from the beamformer was injected into the secondary sources.

The studies from above can be improved regarding the following three aspects:

1. **Control effort** Current studies require canceling both noise and the desired sound first, and then reproducing this desired sound again. Thus, the control effort can be significant.

2. **ANC performance** When an adaptive ANC system needs to reduce both noise and the desired sound, the attenuation of the noise may be degraded because the desired

sounds, e.g., speech and music, are often non-stationary, which can drive the adaptive system to sub-optimal states.

3. **Distortion to the desired sound** When the desired sound needs to be reconstructed and reproduced acoustically, it introduces latency and is prone to different types of distortions, e.g., undesired frequency shaping or binaural cue distortion for ear-level devices.

These aspects lead to the question, "Instead of minimizing the noise plus the desired sound, and then adding the desired sound once again, can the system control only the noise but not the desired sound?"

This article, based on the previous work (Xu and Miller, 2019), proposes a spatially selective ANC system that only controls the unwanted noise while leaving the desired physical sound unaltered. The proposed algorithm is derived based on the hybrid ANC architecture when both the reference and the error signals can be used as inputs for both ANC and the spatial constraint at the same time to optimize the performance. The constraint vector describing the frequency response of the signal from the desired DOA at the error microphone will make sure the desired signal component is physically preserved rather than reconstructed to avoid the aforementioned three issues.

II. PROPOSED SPATIALLY SELECTIVE ANC SYSTEMS

A. Signal definition and cost function

Without losing generality, we assume there is one desired sound source and multiple noise sources in the scene (see Fig. 1). The ANC system contains a total of K microphones, one of which serves as an error microphone. One error microphone with one secondary source is used in the derivation hereinafter for notation brevity, but it can be easily extended to cases with multiple sources and/or microphones. The disturbance signal, $d(n)$, at the error microphone with ANC disabled includes two components,

$$d(n) = s(n) + v(n), \tag{1}$$

where $s(n)$ and $v(n)$ are the desired signal and the noise signal, respectively, and n denotes the time index. We assume that the two signals come from different locations and there is one desired source for now.

We write the error signal $e(n)$ of the hybrid ANC system in *matrix multiplication form* as (Hansen *et al.*, 2012)

$$e(n) = d(n) + \mathbf{w}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{x}(n) \tag{2a}$$

$$= \tilde{\delta}^{\mathrm{T}}\mathbf{x}(n) + \mathbf{w}^{\mathrm{T}}\mathbf{G}^{\mathrm{T}}\mathbf{x}(n) \tag{2b}$$

$$= \mathbf{u}^{\mathrm{T}}\mathbf{x}(n), \tag{2c}$$

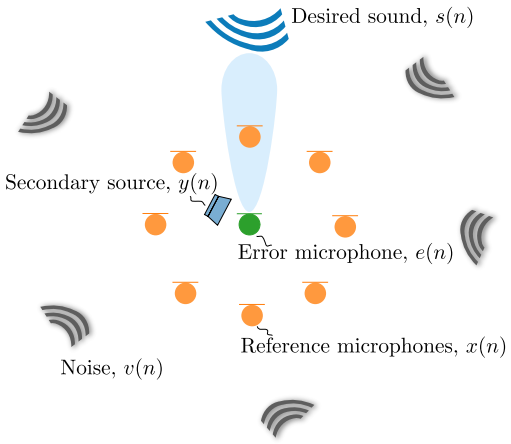


FIG. 1. Concept diagram of a spatially selective ANC system *preserving* the desired sound from one direction and using a secondary source to control noise from other directions at the error microphone (one shown here for demonstration).

where

$$\mathbf{w} \in \mathbb{R}^{KL} = \left[\mathbf{w}_1^{\mathrm{T}}\ \mathbf{w}_2^{\mathrm{T}}\ \ldots\ \mathbf{w}_K^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{3a}$$

$$\mathbf{w}_k \in \mathbb{R}^{L} = \left[w_{k0}\ w_{k1}\ \ldots\ w_{k(L-1)}\right]^{\mathrm{T}}, \tag{3b}$$

$$\mathbf{G} \in \mathbb{R}^{KL\times KL} = \mathrm{diag}\left(\hat{\mathbf{G}}\ \hat{\mathbf{G}}\ \ldots\ \hat{\mathbf{G}}\right), \tag{3c}$$

$$\hat{\mathbf{G}} \in \mathbb{R}^{L\times L} = \begin{bmatrix} \hat{g}_0 & 0 & 0 & \cdots & 0 \\ \hat{g}_1 & \hat{g}_0 & 0 & \cdots & 0 \\ \hat{g}_2 & \hat{g}_1 & \hat{g}_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{g}_{L-1} & \hat{g}_{L-2} & \hat{g}_{L-3} & \cdots & \hat{g}_0 \end{bmatrix}, \tag{3d}$$

$$\mathbf{x}(n) \in \mathbb{R}^{KL} = \left[\mathbf{x}_1^{\mathrm{T}}(n)\ \ldots\ \mathbf{x}_{K-1}^{\mathrm{T}}(n)\ \hat{\mathbf{d}}^{\mathrm{T}}(n)\right]^{\mathrm{T}}, \tag{3e}$$

$$\mathbf{x}_k(n) \in \mathbb{R}^{L} = \left[x_k(n)\ x_k(n-1)\ \ldots\ x_k(n-L+1)\right]^{\mathrm{T}}, \tag{3f}$$

$$\hat{\mathbf{d}}(n) \in \mathbb{R}^{L} = \left[\hat{d}(n)\ \hat{d}(n-1)\ \ldots\ \hat{d}(n-L+1)\right]^{\mathrm{T}}, \tag{3g}$$

$$\tilde{\delta} \in \mathbb{R}^{KL} = \begin{bmatrix} \mathbf{0}^{\mathrm{T}}\ \ldots\ \mathbf{0}^{\mathrm{T}}\ \delta^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}, \tag{3h}$$

$$\mathbf{u} \in \mathbb{R}^{KL} = \tilde{\delta} + \mathbf{G}\mathbf{w}, \tag{3i}$$

where $\mathbf{w}_k$ is the control filter with L length for the k-th channel ($k = 1, 2, \ldots, K$). Superscript $(\cdot)^{\mathrm{T}}$ denotes the transpose, and hat accent $\hat{}$ represents the estimated value. $\hat{\mathbf{G}}$ is the Toeplitz matrix of $\hat{\mathbf{g}}$, which is the estimation of the secondary path impulse response.

We assume that it agrees well with the actual one, i.e., $\hat{\mathbf{g}} = [\hat{g}_0\ \hat{g}_1\ \ldots\ \hat{g}_{L-1}]^{\mathrm{T}} \approx \mathbf{g}$. This assumption holds in many situations and will be made hereinafter. $\delta = [1\ 0\ \ldots\ 0]^{\mathrm{T}}$ is the Dirac delta function. $\hat{\mathbf{d}}(n)$ is the vector of the estimated past values of the disturbance signal $\mathbf{d}(n)$. At the time $n+1$, and after when the ANC is enabled, it is recovered from the error signal $e(n)$ as

$$\hat{d}(n+1) = e(n) - \hat{\mathbf{g}}^{\mathrm{T}}\mathbf{y}(n), \tag{4}$$

where $y(n)$ is the secondary source signal (Kuo and Morgan, 1996).

The spatial selection functionality is achieved by applying a spatial constraint on the cost function of a traditional ANC system. From Eq. (2), the spatial constraint for a single desired source can be expressed as

$$\mathbf{H}^{\mathrm{T}}\mathbf{u} = \mathbf{f}, \tag{5}$$

from the Frost algorithm (Frost, 1972). Matrix $\mathbf{H}$ consists of the relative impulse responses (ReIRs) of the array, i.e.,

$$\mathbf{H} \in \mathbb{R}^{KL\times L} = [\mathbf{H}_1\ \mathbf{H}_2\ \ldots\ \mathbf{H}_K]^{\mathrm{T}}, \tag{6}$$

where $\mathbf{H}_k$ is the Toeplitz matrix from $\mathbf{h}_k = [h_{k0}\ h_{k1}\ h_{k2}\ \ldots\ h_{k(L-1)}]^{\mathrm{T}}$, which is the ReIR between the k-th microphone and a chosen reference microphone (the one closest to the desired source). Vector $\mathbf{f} \in \mathbb{R}^L$ is a constraint vector that describes the frequency response of the signal from the desired direction at the error microphone. It is defined as

$$\mathbf{f} = \mathbf{h}_K = [h_{K0}\ h_{K1}\ \ldots\ h_{K(L-1)}]^{\mathrm{T}}. \tag{7}$$

With ANC enabled, compared with the disturbance signal in Eq. (1), the error signal will also contain two components,

$$e(n) = e_s(n) + v_{\text{ANC}}(n), \tag{8}$$

where $e_s(n)$ is the residual desired signal and $v_{\text{ANC}}(n)$ is the residual noise. By choosing the appropriate reference channel for $\mathbf{H}$ and the vector $\mathbf{f}$ that corresponds to the direction of interest, the proposed system will reduce only the noise, $v(n)$, to $v_{\text{ANC}}(n)$. The constraint vector $\mathbf{f}$ in Eq. (7) can enable the system to preserve the original desired sound component at the error microphone, i.e., $e_s(n) = s(n)$, instead of reconstructing it as in (Dalga and Doclo, 2011; Patel *et al.*, 2020; Serizel *et al.*, 2010).

Finally, the cost function of the proposed system can be written as

$$\begin{aligned} \min_{\mathbf{w}} \; E\left\{e^2(n)\right\} &= \min_{\mathbf{w}} \; E\left\{\mathbf{u}^{\mathrm{T}}\Phi_{\mathbf{xx}}\mathbf{u}\right\} \\ &= \min_{\mathbf{w}} \; E\left\{(\tilde{\delta} + \mathbf{G}\mathbf{w})^{\mathrm{T}}\Phi_{\mathbf{xx}}(\tilde{\delta} + \mathbf{G}\mathbf{w})\right\} \\ &\quad \text{s.t. } \mathbf{H}^{\mathrm{T}}(\tilde{\delta} + \mathbf{G}\mathbf{w}) = \mathbf{f}, \end{aligned} \tag{9}$$

where $\Phi_{\mathbf{xx}} \in \mathbb{R}^{KL\times KL} = E\left\{\mathbf{x}(n)\mathbf{x}^{\mathrm{T}}(n)\right\}$ is the autocorrelation matrix of the tap-stacked input vector, operator $E\{\cdot\}$ denotes mathematical expectation.

B. Optimal solution

The optimal solution can be found by setting the gradient of the cost function to zero and solving the Lagrange multipliers $\lambda \in \mathbb{R}^L$ (Haykin, 2002). The derivation of such an optimal solution can be found in appendix A, where the solution is provided in Eq. (A4). It

can be re-written in the following form for interpretation,

$$\begin{aligned}\mathbf{w}_{\text{opt}} = &- \Phi_{\mathbf{rr}}^{-1}\phi_{\mathbf{r}d} \\ &+ \Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{G}\Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H} + \rho\mathbf{I}\right)^{-1}\mathbf{f} \\ &- \Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{G}\Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H} + \rho\mathbf{I}\right)^{-1} \\ &\qquad\left(\mathbf{H}^{\mathrm{T}}\tilde{\delta} - \mathbf{H}^{\mathrm{T}}\mathbf{G}\Phi_{\mathbf{rr}}^{-1}\phi_{\mathbf{r}d}\right),\end{aligned} \tag{10}$$

where

$$\mathbf{r}(n) = \mathbf{G}^{\mathrm{T}}\mathbf{x}(n), \tag{11a}$$

$$\Phi_{\mathbf{rr}} = E\left\{\mathbf{r}(n)\mathbf{r}^{\mathrm{T}}(n)\right\} + \beta\mathbf{I}, \tag{11b}$$

$$\phi_{\mathbf{r}d} = E\{\mathbf{r}(n)d(n)\}. \tag{11c}$$

This optimal solution has three terms. The first term is the Wiener solution of a hybrid ANC system controlling all the observable sounds. The second term is due to the spatial constraint, which is similar to the beamformer solution in (Frost, 1972), but the secondary path matrix $\mathbf{G}$ is added due to the physical constraint of the ANC system. The third term provides the coupling of the two subsystems. All three terms contribute to calculating the control filter $\mathbf{w}$ such that only the noise from the undesired directions is minimized. The residual desired signal component in the residual error signal after control is the original desired physical sound left unaltered by the proposed system. Further details about preserving or reconstructing this desired physical sound will be emphasized in Section IV.

Note that matrix $\mathbf{H}^{\mathrm{T}}\mathbf{G}\Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H}$ is rank deficient due to $\mathbf{G}$ being rank deficient. (There are delays in the secondary paths.) Thus, a Tikhonov regularization factor ρ has been applied to the diagonal elements to make it invertible (Tikhonov and Arsenin, 1977).

Note that Eq. (11b) has also been added with a regularization factor β compared to Eq. (A4), which is equivalent to adding a penalty term, the l_2 norm of the control filter $||\mathbf{w}||_2$ in Eq. (9). This is essentially the leaky version of the algorithm for a more robust ANC system (Cartes *et al.*, 2002; Elliott *et al.*, 1992). The robustness of the system will be further discussed in Section II E.

C. Adaptive solution

Commonly, adaptive algorithms are used to reduce computations and handle fast environmental changes. The derivations of the adaptive solution in the proposed method can be found in appendix B. The final solution is expressed as

$$\mathbf{w}(0) = \mathbf{q}, \tag{12a}$$

$$\mathbf{w}(n+1) = \mathbf{P}\left[\mathbf{w}(n) - \mu\mathbf{G}^{\mathrm{T}}\mathbf{x}(n)e(n)\right] + \mathbf{q}, \tag{12b}$$

where

$$\mathbf{P} = \mathbf{I} - \mathbf{G}^{\mathrm{T}}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{H} + \gamma\mathbf{I}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{G}, \tag{13a}$$

$$\mathbf{q} = \mathbf{G}^{\mathrm{T}}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{H} + \gamma\mathbf{I}\right)^{-1}\left(\mathbf{f} - \mathbf{H}^{\mathrm{T}}\tilde{\delta}\right), \tag{13b}$$

and μ is the step-size. Fig. 2 illustrates the block diagram of the proposed adaptive spatially selective ANC system.

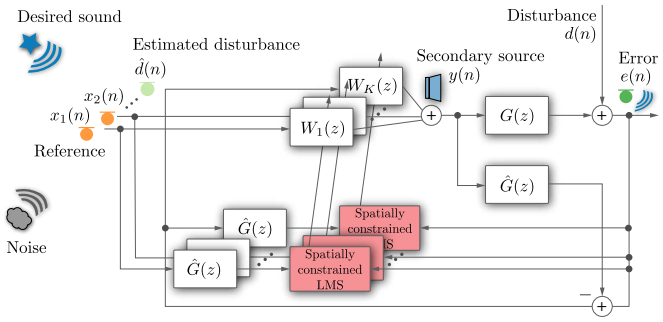


FIG. 2. Block diagram of proposed spatially selective ANC system. The adaptive hybrid ANC algorithm is spatially constrained.

Notice that the solution is coupled by the adaptive algorithm in the ANC subsystem and the adaptive Frost algorithm. The spatial constraint in Eq. (9) resulted in $\mathbf{P}$ and $\mathbf{q}$. Without it, the solution would become $\mathbf{w}(n+1) = \mathbf{w}(n) - \mu\mathbf{G}^{\mathrm{T}}\mathbf{x}(n)e(n)$, which is a traditional adaptive hybrid ANC solution minimizing the overall error signal (Elliott, 2000; Hansen *et al.*, 2012).

Notice that a regularization factor γ in Eq. (13) has been added compared to Eq. (B5). Similarly to the optimal solution, although the matrix $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is invertible due to the use of ReIRs, that is, there is always an identity matrix in $\mathbf{H}$, matrix $\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{H}$ is rank-deficient. Therefore, γ has been added for inversion.

D. Spectral weighting

In some cases, spatial filtering alone may not be sufficient due to limitations of the array, e.g., filter length, array configuration. To deal with this, the proposed method can be further

improved by applying a spectral weighting filter, i.e.,

$$\mathbb{F} \in \mathbb{R}^L = \mathbf{S}\mathbf{h}_K \tag{14}$$

instead of $\mathbf{f}$ in Eq. (7). $\mathbf{S} \in \mathbb{R}^{L \times L}$ is the Toeplitz matrix of the impulse response of the spectral weighting filter to attenuate the frequency range that is not of interest. Note that $\mathbf{S}$ is a digital filter, which can be designed to be minimum-phase. A non-minimum-phase filter is not desired since it leads to delays in the error signal.

This technique is performed provided that the frequency range attenuated by the spectral filter has little overlap with that of the desired signal. Otherwise, one needs to consider the trade-off between noise reduction and signal distortion.

E. Robustness

Robustness is an important factor to consider. The robustness issues in ANC systems can be contributed by non-stationary inputs, low SNR, and/or secondary path changes (Cartes *et al.*, 2002; Elliott, 2000). Many beamforming systems are subject to numerical and/or spatial robustness issues arising from signal mismatches, which are due to mismatches between the presumed and actual relative transfer functions (Gannot *et al.*, 2017; Vorobyov, 2013). We will mainly focus on the numerical robustness issue due to signal mismatches herein.

For the proposed joint optimization problem, the robustness of the system can be from the following aspects:

1. Robustness of the ANC subsystem

2. Robustness of the beamforming subsystem

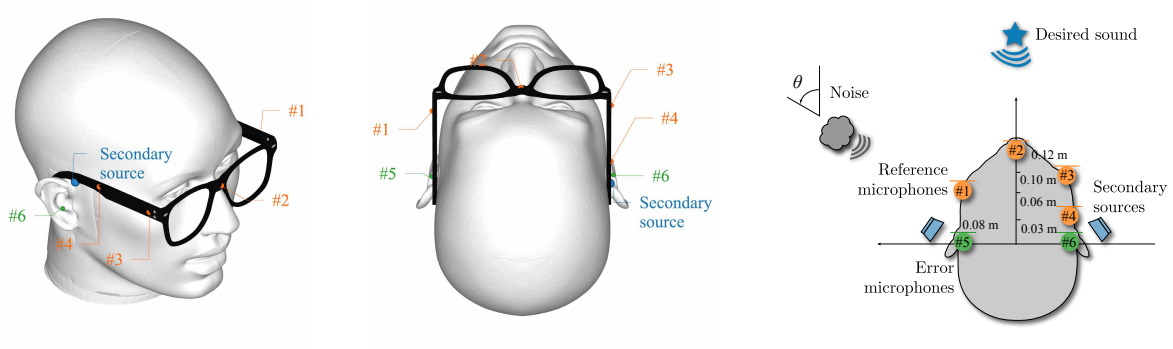
3. ANC affects the beamforming subsystem

4. Beamforming affects the ANC subsystem (due to signal mismatches)

The former two aspects can be easily understood since each subsystem can inherently have its own robustness issues. The solutions to these issues can also be easily found in many studies. For example, the ANC subsystem can use a leaky algorithm constraining $||\mathbf{w}||_2$ in the cost function (Cartes *et al.*, 2002; Tobias and Seara, 2004). As for a robust beamformer, the diagonal loading method can be used. This is achieved by constraining the white noise gain (WNG) $||\mathbf{u}||_2$ in the cost function (Cox *et al.*, 1987; Li *et al.*, 2003; Vorobyov, 2013).

The latter two issues are due to the joint optimization of ANC and beamforming. The third issue arises from the secondary paths in ANC systems, particularly the delays from the acoustic propagation and the electronics. They can be coupled with the beamforming constraints resulting in an adverse effect, e.g., the rank deficiency of $\mathbf{G}$ as discussed previously. Regularization factors ρ and γ have been added to solve the problem (Hansen, 2010; Tikhonov and Arsenin, 1977).

As for the fourth aspect, the beamforming constraint can be unstable due to signal mismatches and thus affect the ANC performance. This is similar to designing a robust beamformer. Constraining the WNG in the cost function can stabilize the system. The optimal solution presented previously has been applied with the regularization factor β for robustness (Cox *et al.*, 1987; Vorobyov, 2013).



(a) (b) (c)

FIG. 3. (a) The isometric view and (b) the top view of a KEMAR manikin with a pair of AR glasses with a six-microphone array. (c) Microphone setup.

III. NUMERICAL SIMULATION ON AUGMENTED REALITY (AR) GLASSES

In this section, we put the proposed algorithm in a pair of open-fitting AR glasses.

A. System setup

A six-channel system from the EasyCom dataset (Donley *et al.*, 2021) is shown in Fig. 3. The four microphones in the frame (labeled as #1 to #4) can be used for the reference microphones. For each ear, the corresponding binaural microphone (labeled as either #5 or #6) was used as the error microphone. The single-ear (right side) system is considered in this article for brevity, but it can be easily extended to a binaural case. The control performance at microphone #6 is focused hereinafter.

It was assumed that the desired speech was at $\theta = 0°$ and the noise source was at $\theta = 60°$. The desired signal was a 20-second male speech (Acclivity, 2006). The noise was a speech babble noise from the NOISEX-92 database (Varga and Steeneken, 1993). The noise level

was adjusted such that the original clean speech was not intelligible when mixed, i.e., the *a priori* SNR was −13.2 dB. The waveforms and the spectrograms of the clean speech and the noisy speech can be seen in Figs. 4a and 4b, respectively. These signals were influenced by the KEMAR manikin (Burkhard and Sachs, 1975) and the glasses. The frequency response of the acoustic secondary path was acquired from the COMSOL Multiphysics software, where the secondary source was simulated as a perfect point sound source and was located at about 0.05 m above the error microphone #6 as shown in Fig. 3a. In this article, we assumed that the sound source was not constrained by transducer characteristics (such as excursion limit and transducer resonance), and the response of the electrical control system and the transducers could be modeled with pure delays. The sampling rate was 48 kHz, and the filter length $L = 768$. The secondary path delay had ten samples (208.3 μs), which included both the acoustic propagation delay and the electronics delay in practice.

B. Performance metrics

To quantitatively evaluate the control performance, the noise reduction (NR) level of the ANC system is defined as

$$\mathrm{NR} = 10\log_{10}\left(\frac{E\left\{v^2(n)\right\}}{E\left\{v_{\mathrm{ANC}}^2(n)\right\}}\right), \tag{15}$$

where the NR level should be as high as possible.

In addition, the speech distortion index (SDI) (Chen *et al.*, 2006) is used to monitor any potential distortion in the residual speech component, and it is calculated as

$$\mathrm{SDI} = 10\log_{10}\left(\frac{E\left\{\left[s(n) - e_s(n)\right]^2\right\}}{E\left\{s^2(n)\right\}}\right), \tag{16}$$

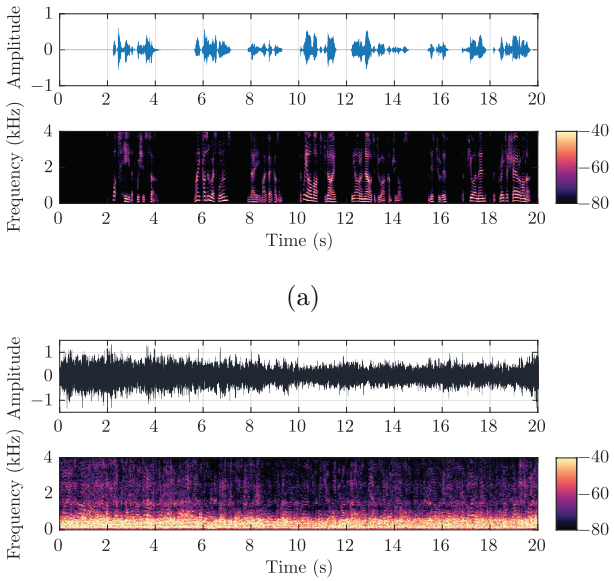


FIG. 4. Waveforms and spectrograms of (a) the desired clean speech and (b) the noisy speech at the error microphone.

where $e_s(n)$ should match $s(n)$ as much as possible, and thus the SDI value should be as small as possible.

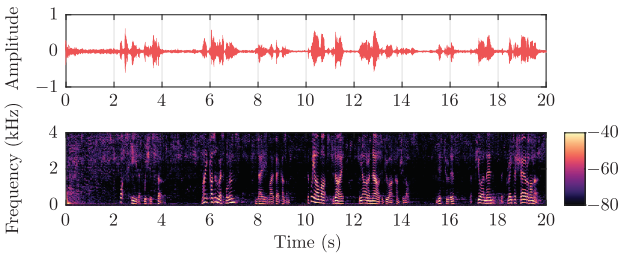
Signals $e_s(n)$ and $v_{\mathrm{ANC}}(n)$ can be decoupled by recording the history of the control filter $\mathbf{w}$, which is then used to re-filter either the desired speech signal or the noise. For example, one can nullify the desired speech signal while maintaining the noise to observe the NR level. Similarly, one can observe the SDI value by nullifying the noise. This method is only used to quantitatively evaluate the components in the error signal in this article. In reality, the desired signal and the noise are unknown. It is typical to estimate them at different time intervals in speech processing and then estimate the NR levels and the SDI values.

C. Control performance

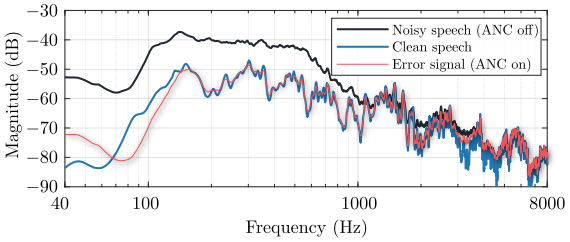
The step-size μ used in the simulation was a variable step-size (VSS) to ensure the system with a fast convergence speed and small misadjustment for the desired speech signal (Aboulnasr and Mayyas, 1997). The parameters chosen for calculating the VSS were: $\mu_{\max} = 0.0001$, $\mu_{\min} = 0.000008$, $\alpha_{\text{VSS}} = 0.99998$, $\gamma_{\text{VSS}} = 0.00001$ and $\beta_{\text{VSS}} = 0.99999$. The regularization factor γ in Eq. (13) was chosen to be 0.0001. A minimum-phase high-pass filter with a cut-off frequency at 140 Hz was applied as the spectral weighting filter discussed in Eq. (14).

The waveform and the spectrogram of the residual error signal are shown in Fig. 5a. After ANC was enabled using the proposed method, the noise component was considerably attenuated within the first two seconds, leaving only the desired speech signal in good agreement with the original clean speech in Fig. 4a overall.

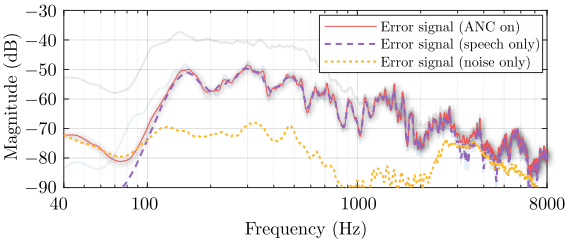
The spectra of the noisy speech, the clean speech and the total residual error signal with ANC enabled are shown in Fig. 5b. Although the spectrum of the error signal followed the majority of the clean speech, there was still some minor residual noise below 100 Hz. As shown in Fig. 5c, decoupling the speech and the noise components as described in Section III B confirms that the system was mainly bound by the ANC subsystem. From the last 10 s of the signals, the SNR has been improved from −13.9 dB to 15.2 dB in total, which enhanced the unintelligible speech significantly. The NR level was 29.1 dB, and the SDI value was shown to be –25.1 dB above 100 Hz. The SDI was low enough for the listener not to notice any undesired distortion.



(a)



(b)



(c)

FIG. 5. (a) Waveform and the spectrogram of the error signal with ANC enabled. (b) Spectra of the noisy speech, the clean speech, the total error signal with ANC enabled, and (c) the decoupled speech and noise components in the error signal. The spectra are from the last 10 s period.

D. Spectral weighting

As discussed in Section II D, the performance of the spatial constraint may be sub-optimal due to either an insufficient number of channels or limited filter lengths. Fig. 6 shows the control performance without the spectral weighting filter. The error signal cannot be controlled adequately below 100 Hz. The reason can be found by performing the hybrid ANC and the Frost algorithms separately with the same six microphones. Although the

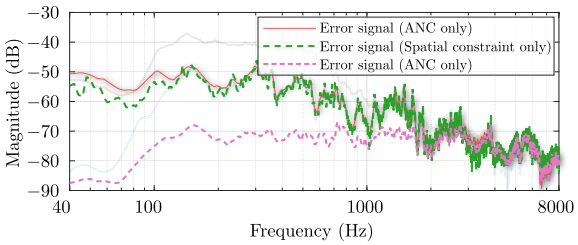
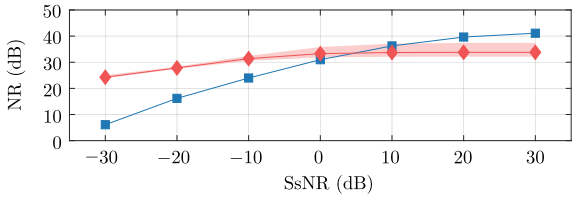


FIG. 6. Spectra of the error signal from the proposed method, the error signal from the spatial constraint only and from the ANC system only after control *without* using the minimum-phase high-pass filter with a cut-off frequency at 140 Hz.

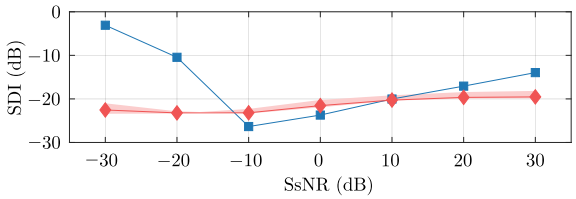
ANC subsystem can reduce the noise across the spectrum using the hybrid control, the spatial constraint cannot below 100 Hz due to the limited length of the filters. Therefore, it is necessary to use the spectral weighting method.

E. Robustness

This part follows the system robustness discussions in Section II E. We mainly examine how sensor noise (signal mismatches) from the beamforming constraint affects the ANC system and the control performance. To examine the robustness of the system, sensor noise was added to some channels, e.g., microphone #1, #3, #4 and #5. It is common to model the sensor noise as Gaussian white noise (Van Trees, 2002) with the power of σ_{n}^2. The signal-to-sensor-noise-ratio (SsNR) at microphone #5 was used to represent different levels of sensor noise. The optimal solution in Eq. (10) was used to calculate the ANC control filter. Here, we show the importance of β and ρ for the robustness of the system.



(a)



(b)

FIG. 7. (a) The NR level and (b) the SDI value of the error signal with respect to different SsNRs. The blue square marks are the results by choosing $\beta = \rho = 10\sigma_{\mathrm{n}}^2$. The red diamond marks are the results by choosing $\beta = \lambda_{\mathrm{max}}^1/10000$ and $\rho = \lambda_{\mathrm{max}}^2/10000$. The shaded areas represent the results with the ratio from 5000 to 50000.

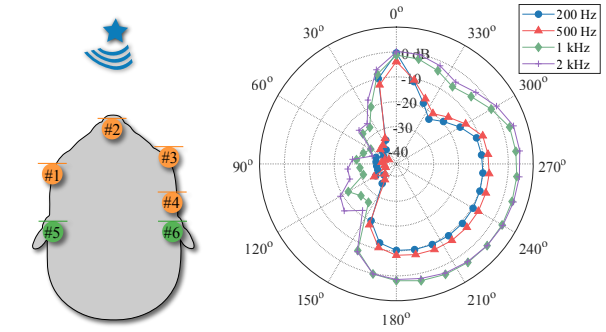
Beamforming problems typically follow the rule of $10\sigma_{\mathrm{n}}^2$ to choose the regularization factor (Li *et al.*, 2003; Shahbazpanahi *et al.*, 2003; Vorobyov *et al.*, 2003). Fig. 7 (blue square mark) shows the NR and SDI results of the proposed ANC system using $\beta = \rho = 10\sigma_{\mathrm{n}}^2$ under different SsNRs. The problem with this method is that, for example when the sensor noise is small, e.g., SsNR = 30 dB, β and ρ are too small to have a regularization effect. Although the noise can be significantly reduced (NR = 41.1 dB), the desired speech is also highly distorted (SDI = –14.0 dB). On the other hand, when the sensor noise is great, e.g., SsNR = –30 dB, β and ρ are very large and can over-regulate the system. The system can neither control noise sufficiently (NR = 6.1 dB), nor retain the original desired speech (SDI = –3.1 dB).

We show that β and ρ can be chosen depending on the largest eigenvalue of the matrix to be inverted. Assume the largest eigenvalues of $E\left\{\mathbf{r}(n)\mathbf{r}^{\mathrm{T}}(n)\right\}$ and $\mathbf{H}^{\mathrm{T}}\mathbf{G}\Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H}$ are $\lambda_{\max}^{1}$ and $\lambda_{\max}^{2}$, respectively. Typically, the ratios between $\lambda_{\max}^{1}$ and β and between $\lambda_{\max}^{2}$ and ρ depend on different systems. Here, we found that the ratio between 5000 and 50000 had good results. The performance under this range is shown in Fig. 7 with shaded areas, and the result with $\beta = \lambda_{\max}^{1}/10000$ and $\rho = \lambda_{\max}^{2}/10000$ is depicted as an example (red diamond mark). It is obvious the system performance has been maintained well across different levels of sensor noise. Particularly when the input signals have been disastrously perturbed, i.e., SsNR = –30 dB, the system could still exhibit good behavior. The NR level had 24.3 dB and the SDI was maintained at –22.5 dB. Thus, by choosing the regularization factors based on the largest eigenvalue instead of the sensor noise power, the proposed system can achieve a good result even with extreme cases of sensor noise.

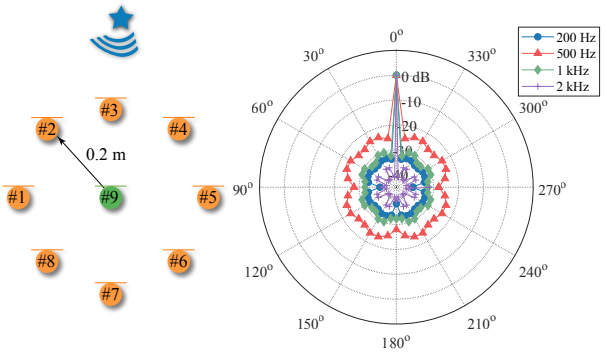
F. Directivity

The direction-dependent NR performance of the demonstrated AR glasses ANC system is shown in Fig. 8a for different frequency bands. In this experiment, the desired sound source is fixed at the 0 degree angle, and a pink noise source is placed at various angles in the horizontal plane. The optimal solutions were calculated.

At the desired direction $\theta = 0°$, all signals were maintained including the noise. The NR performance was the best at $\theta \in (30°,\ 150°)$. The noise could be reduced by at least 20 dB. Low frequencies, e.g., below 500 Hz, had more than 30 dB reduction. The system had an unsatisfactory performance at $\theta \in (150°,\ 300°)$. This was due to the ANC capability



(a)



(b)

FIG. 8. Directivity plot of the residual noise at the error microphone #6 for a single noise source by (a) the demonstrated AR glasses configuration and (b) a system with eight reference microphones in a circular formation with an error microphone #9 at the center for the pink noise.

of the specific microphone array configuration. In these directions, the noise reached the error microphone first and then the reference microphones. Thus, the system causality was violated and only the feedback subsystem was in operation, which was only effective below 500 Hz with about 10 dB reduction. High frequencies were slightly increased due to the waterbed effect (Skogestad and Postlethwaite, 2005).

One of the potential solutions is to decrease the delay in the secondary path, which mainly depends on the electronic components in the device. Another way is to adjust the array configuration. As shown in Fig. 8b, eight reference microphones are in a circular formation

with an error microphone at the center. Although ideal, it demonstrated that it is possible to cancel noise in every direction except for the desired direction if the causality of the ANC subsystem can be maintained. Depending on the specific application, the array design will change, and the directivity pattern will change accordingly.

IV. COMPARISON WITH EXISTING METHODS

One unique feature of the proposed method is to truly preserve the original desired physical sound rather than reconstruct it. The results from the proposed method will be compared with the ones from the previous works (Dalga and Doclo, 2011; Patel *et al.*, 2020; Serizel *et al.*, 2010).

We would like to highlight the fact that the previous works had been developed for different systems with different ANC and beamforming algorithms. For ANC, Serizel *et al.* (2010) and Patel *et al.* (2020) used the feedforward configuration, while Dalga and Doclo (2011) used the hybrid control. For the beamforming, Serizel *et al.* (2010) and Dalga and Doclo (2011) used the MWF, while Patel *et al.* (2020) used the superdirective beamformer. For consistency in this article, we evaluated all the methods for the aforementioned open-fitting AR glasses with the fixed six-array microphones. The monaural setup at microphone #6 is still generally considered for simplicity of discussion. For a fair comparison, the hybrid ANC control and the Frost algorithm were used for all the cases. The optimal solutions were also computed for most cases, except Section IV B. *The main goal for the comparison was to see how the desired signal was obtained (e.g., reconstructed or preserved) in these systems, and their implications.*

The three configurations were configured as follows:

1. For (Dalga and Doclo, 2011; Serizel *et al.*, 2010), the ANC and beamformer modules can be *partially coupled*. Microphones #1 to #6 were used for ANC, and two microphones (#2 and #4) were also used as the beamforming array. The extracted signal from the beamformer had 1 ms delay, and the gain was 0 dB as provided by Serizel *et al.* (2010). The extracted signal was added to the error signal [see Fig. 5 in (Serizel *et al.*, 2010)].

2. For the *decoupled* configuration by Patel *et al.* (2020), microphones #1, #3 were used as the reference microphones for the ANC subsystem, and microphones #2 and #4 were used as the beamforming array. The other error microphone #5 was not available to control #6. Note that such a decoupled configuration requires dedicated reference microphones with error microphones for ANC (microphone #1 with #5 for the left channel ANC, #3 with #6 for the right). Thus, only microphones #2 and #4 are left for the beamformer. The extracted signal from the beamformer had 5 ms delay. The extracted signal was added to the secondary source [see Fig. 3 in (Patel *et al.*, 2020)].

3. The proposed method, as described in the previous section.

The aforementioned three aspects - control effort of the secondary source, control performance for multiple noise sources and binaural localization cues and latency of the desired sound - are discussed below.

A. Control effort

Fig. 9a shows the secondary source signal $y(n)$ in the three systems when the desired speech was at 0°, the pink noise was at 60°, and the *a priori* SNR was 0 dB. It is clear that the secondary source signals in the partially coupled and the decoupled configurations contained the desired speech. These systems needed to cancel both the desired speech and noise first and then reconstruct and reproduce the desired speech. It is even more so for the decoupled configuration since the desired speech needs to be directly injected into the secondary source signals [see Fig. 3 in (Patel *et al.*, 2020)]. On the contrary, the one from the proposed method was only for the noise, indicating it controlled the noise only and thus preserving the original desired speech.

Another way for confirmation is to compare the control effort for various *a priori* SNRs. We used the secondary source energy $E\{y^2(n)\}$ instead of $||\mathbf{w}||_2^2$ since the former provides a clearer physical representation (Elliott, 2000, p. 147). The relative energy consumption $\mathscr{E}_y$ in percentage was calculated as

$$\mathscr{E}_y = \frac{E\{y^2(n)\}}{E\{y_{\text{ref}}^2(n)\}} \times 100\%, \tag{17}$$

where $E\{y_{\text{ref}}^2(n)\}$ is the mean-square of the secondary source signal in (Dalga and Doclo, 2011; Serizel *et al.*, 2010) for a time period (e.g., 20 seconds) as the reference.

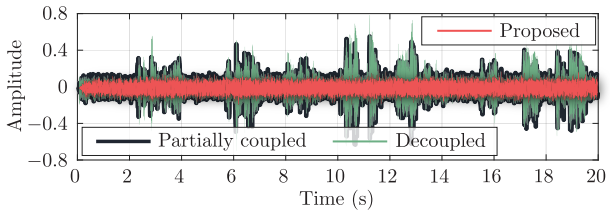
The energy of the secondary source and the corresponding NR levels are shown in Figs. 9b and 9c, respectively. When the noise level was high, e.g., SNR $= -15$ dB, the energy difference was small since most energy was devoted to controlling the noise for all cases. The NR levels were also similar. However, as the desired speech became more and more

prominent, the difference became more and more apparent. For example, when SNR = 10 dB, the environment is relatively quiet. The other two configurations still used a vast amount of energy. Particularly, the decoupled configuration used 50% more energy than the partially coupled due to better control performance. (The reason for this will be discussed in the next subsection.) These systems see the desired speech as "noise" too and attempt to cancel it first, and then try to reconstruct the desired speech once again.

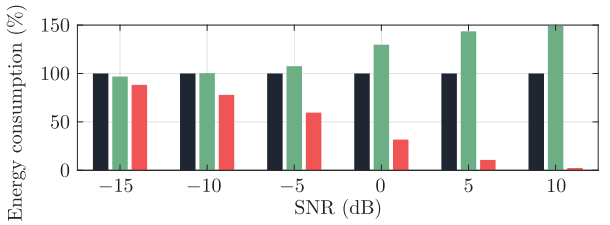
On the other hand, the proposed algorithm barely needed to change the original desired signal as illustrated in the time domain in Fig. 9a. It required only 2% energy while yet achieving a better NR level than the partially coupled configuration. These results further confirm that the proposed system only reduces the noise and truly *preserves* the desired sound.

B. Noise attenuation for multiple noise sources

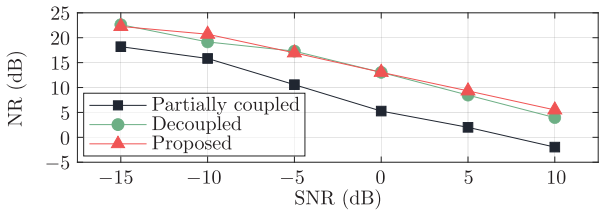
The ANC performance comparison of the three configurations has been partially demonstrated for a single noise in Fig. 9c. A more practical situation is that the noises are from multiple directions. In this case, the uncorrelated pink noise was set coming from five directions $60°, 90°, 120°, 300°$ and $330°$ with the same level, while the DOA of the desired speech remained to be from $0°$ and other configurations were kept the same. Fig. 10a shows the overall error signal. The proposed system had the closest result to the desired clean speech. The noise and the speech components in the error signal can be decoupled for further observations.



(a)



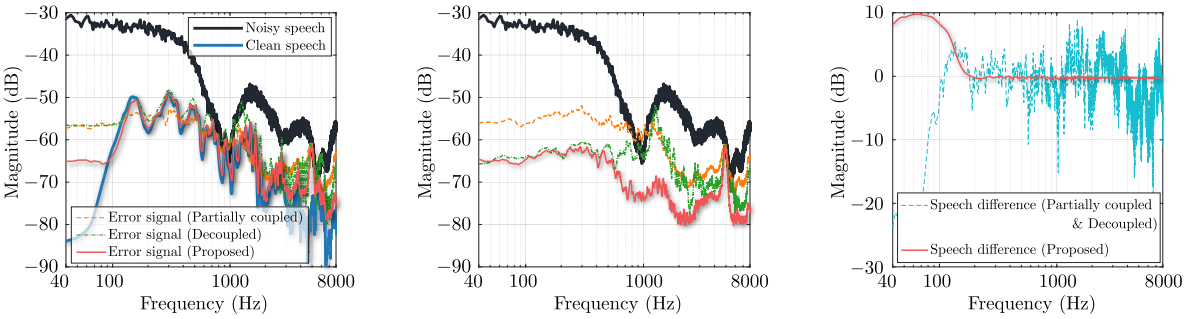
(b)



(c)

FIG. 9. (a) Secondary source signals from the three configurations when the *a priori* SNR = 0 dB. (b) Relative secondary source energy consumption, and (c) the NR levels for different *a priori* SNRs.

The noise components in the error signal are shown in Fig. 10b. For the other two configurations, which tried to cancel both the desired speech and noise, more microphones (in the partially coupled configuration) can lead to a worse performance. The desired speech in the input signal can result in a greater eigenvalue spread of the correlation matrix of the input signal, thus limiting the step-size and causing a slower adaption speed (Haykin, 2002). The maximum step-size before making the system diverge for the partially coupled system was 0.00004, whereas the one for the decoupled was 0.00008. Thus, configuration 2 had a

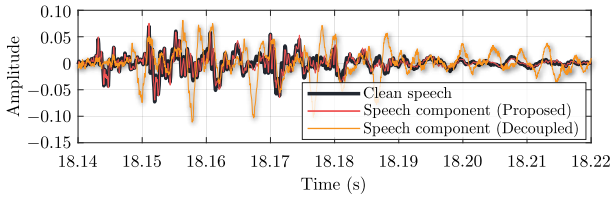


(a) Overall (b) Noise component (c) Speech component difference

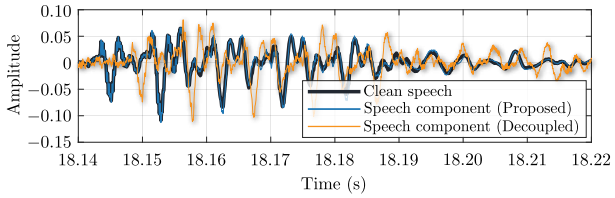
FIG. 10. Spectra of the noisy speech, the clean speech, (a) the overall error signals, (b) the noise components in the error signals and (c) the speech component differences compared to the clean speech in the error signals with ANC enabled in the three configurations. The spectra are from the last 10 s period.

better ANC performance in both Figs. 9c and 10b. The proposed configuration did not need to cancel the speech. Thus, it has the best NR performance.

The speech component differences compared to the original clean speech are shown in Fig. 10c. The other two systems used the same array for signal extraction, thus having the same result. The reconstructed desired speech from these systems had about ±10 dB difference compared to the clean speech in general. On the other hand, the proposed system did not have this issue since it left the original desired physical sound unaltered. Thus, the difference above 140 Hz was essentially zero. The difference below 140 Hz was due to the spectral weighting filter as discussed previously. This range did not cause any noticeable auditory distortion.



(a) Left ear (microphone #5)



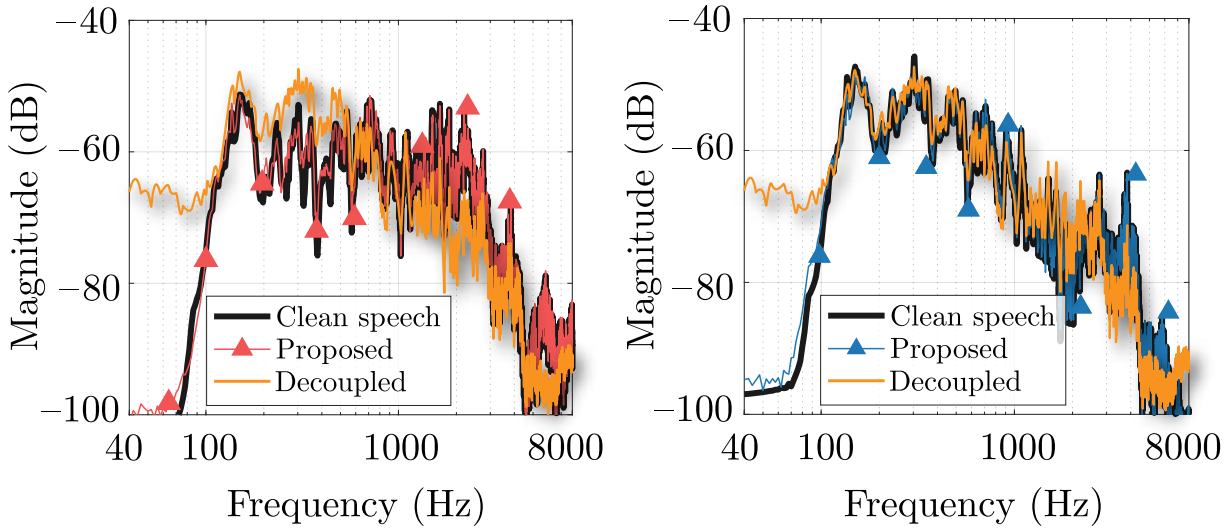
(b) Right ear (microphone #6)

FIG. 11. Waveforms of the clean speech at the error microphone and the speech components in the error signals from the decoupled and the proposed systems at two ears. The desired sound is at 60° and the noise comes from 0°.

C. Binaural localization cues and latency

When the desired sound comes from other directions than 0°, binaural localization cues should be enabled for the user. The DOA of the desired speech was moved to 60°, whereas the pink noise came from 0° instead.

Fig. 11 shows the waveforms of the speech components in the decoupled and the proposed systems compared to the clean speech at two ears. The result in the partially coupled system was the same as the decoupled but with 1 ms delay instead of 5 ms. The partially coupled and the decoupled systems had the reconstructed desired signals at microphone #4. Unless a binaural beamformer was used, the binaural localization cues were lost. This is also shown in Fig. 12 in the frequency domain. When the DOA of the desired source is at 60°, the spectral property of the reconstructed desired signal at microphone #4 may be similar



(a) Left ear (microphone #5)

(b) Right ear (microphone #6)

FIG. 12. Spectra of the clean speech at the error microphone and the speech components in the error signals from the decoupled and the proposed systems at two ears. The desired sound is at 60° and the noise comes from 0°.

to that of microphone #6 due to their close proximity, though only between 100 Hz and 2 kHz as shown in Fig. 12b. For the left side, the reconstructed signal is monaural and still at microphone #4, which is significantly different than that of microphone #5 across the spectrum as shown in Fig. 12a.

On the other hand, it is apparent that the speech component in the error signal from the proposed system agreed with the original clean speech very well, in both time and frequency domains. This once again confirms that the proposed system can preserve the natural binaural localization cues since it left the desired physical sound unaltered.

V. REMARKS & FUTURE DIRECTIONS

For the purpose of brevity, it was assumed that there was one desired source throughout this article. For multiple desired sources from different directions, multiple spatial constraints can be added to the cost function in Eq. (9), i.e., $\mathbf{H}_{s1}^{T}(\tilde{\delta}+\mathbf{G}\mathbf{w}) = \mathbf{f}_{s1}, \mathbf{H}_{s2}^{T}(\tilde{\delta}+\mathbf{G}\mathbf{w}) = \mathbf{f}_{s2}, \ldots$ pointing to the directions of the desired sources. These constraints can also be combined into one. More information can be found in the linear constraint minimum variance beamforming algorithm (Van Trees, 2002; Van Veen and Buckley, 1988).

The desired sound and the noise can be from the same direction. This can happen when either both sources are located in the same direction, or the noise source is in another direction but the reverberation of the noise is mixed with the desired source. This problem requires further investigation. Possible solutions may be found in the field of *Blind Source Separation*. Studies such as (Mukai *et al.*, 2006) can separate mixed signals even when two sources come from the same direction.

The demonstrated system was a pair of open-fitting AR glasses, which allows the system to be fully coupled, that is, all the microphones to be used for both ANC and the spatial constraint. Another ANC application, for example, a pair of close-fitting ANC headphones, may not be able to use error microphones in the earcups for the spatial constraint since the disturbance signals are highly attenuated. Using all the microphones may not provide any significant improvements. Therefore, using a partially coupled system can reduce computation.

The secondary sources in such a personal AR glasses system are miniaturized speakers. The proposed system only needs to cancel the noise, so the transducer excursion limit can be relaxed. This is favorable in acoustic transducer designs. However, the response of miniaturized speakers can still be sub-optimal at some frequencies, which can affect the system control performance. Therefore, the secondary path in this article was from the COMSOL simulation with an ideal sound source to eliminate this factor for the purpose of brevity. Other applications, such as ANC headrests or ANC windows, which can also adopt the proposed algorithm, may not have this issue since the loudspeakers are not necessarily required to be miniaturized.

The acoustic feedback from the secondary sources to reference microphones can affect the system significantly in practice. It should be considered in the future.

VI. CONCLUSION

In this article, we have presented a spatially selective ANC system that truly preserves the desired sound rather than canceling it and then reconstructing it again. The Frost spatial constraint was imposed on the cost function of a traditional hybrid ANC system, and both the optimal and the adaptive solutions were derived. The method was examined in a pair of AR glasses with six microphones. Overall, the system exhibited good performance in controlling noise coming from undesired directions. The SNR was improved from -13.9 dB to 15.2 dB while the SDI was kept at -25.1 dB. The system could also maintain good robustness when it was disastrously perturbed. Even when the signal mismatch level was 30 dB higher than that of the desired signal, by choosing the regularization factor based

on the largest eigenvalue, the noise could still be controlled by 24.3 dB and the SDI was maintained at −22.5 dB. However, the directivity was bound by the causality of the ANC system, which mainly depended on the array configuration.

Compared to the state-of-the-art systems, the desired sound in the proposed system was indeed preserved while noise from other directions was minimized. The proposed system used much less control effort while still achieving the best ANC performance since it controlled only the noise instead of noise plus speech. Furthermore, when the original desired speech at the error microphone is preserved, there is no need to reconstruct the natural binaural localization cues as in other systems. Future work includes considering acoustic feedback control and examining the system in a non-ideal environment, e.g., in a reverberant room.

APPENDIX A: OPTIMAL SOLUTION DERIVATION

By taking the gradient of Eq. (9) and using the chain rule (Petersen and Pedersen, 2012), it is found that

$$\nabla_{\mathrm{w}} J = \frac{\partial J}{\partial \mathbf{w}} = \frac{\partial \mathbf{u}}{\partial \mathbf{w}} \frac{\partial J}{\partial \mathbf{u}} = \mathbf{G}^{\mathrm{T}} \left(\Phi_{\mathbf{xx}} \mathbf{u} + \mathbf{H} \lambda \right). \tag{A1}$$

Setting the gradient of the cost function to zero, the optimal solution can be written as

$$\begin{aligned} \mathbf{w}_{\mathrm{opt}} &= -(\mathbf{G}^{\mathrm{T}} \Phi_{\mathbf{xx}} \mathbf{G})^{-1} \mathbf{G}^{\mathrm{T}} (\Phi_{\mathbf{xx}} \tilde{\delta} + \mathbf{H} \lambda) \\ &= -\left(\Phi_{\mathbf{rr}}^{-1} \phi_{\mathbf{r}d} + \Phi_{\mathbf{rr}}^{-1} \mathbf{G}^{\mathrm{T}} \mathbf{H} \lambda \right), \end{aligned} \tag{A2}$$

where $\mathbf{r}(n) = \mathbf{G}^{\mathrm{T}} \mathbf{x}(n)$, $\Phi_{\mathbf{rr}} = E\left\{ \mathbf{r}(n) \mathbf{r}^{\mathrm{T}}(n) \right\}$ and $\phi_{\mathbf{r}d} = E\{\mathbf{r}(n) d(n)\}$. The Lagrangian λ can be found by putting Eq. (A2) in the spatial constraint $\mathbf{H}^{\mathrm{T}}(\mathbf{G}\mathbf{w}_{\mathrm{opt}} + \tilde{\delta}) = \mathbf{f}$,

$$\lambda = \left(\mathbf{H}^{\mathrm{T}} \mathbf{G} \Phi_{\mathbf{rr}}^{-1} \mathbf{G}^{\mathrm{T}} \mathbf{H} \right)^{\dagger} \left[(\mathbf{H}^{\mathrm{T}} \tilde{\delta} - \mathbf{f}) - \mathbf{H}^{\mathrm{T}} \mathbf{G} \Phi_{\mathbf{rr}}^{-1} \phi_{\mathbf{r}d} \right], \tag{A3}$$

where superscript $(\cdot)^{\dagger}$ denotes pseudoinverse. Matrix $\mathbf{G}$ is rank deficient due to the delays in the secondary paths in the ANC system. Note that the autocorrelation matrix of the filtered reference signals $\Phi_{\mathbf{rr}}$ is typically full rank and thus is invertible.

Using Eqs. (A2) and (A3), the optimal solution can be found as

$$\begin{aligned}\mathbf{w}_{\text{opt}} = &- \left[\mathbf{I} - \Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{G}\Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H}\right)^{\dagger}\mathbf{H}^{\mathrm{T}}\mathbf{G}\right]\Phi_{\mathbf{rr}}^{-1}\phi_{\mathbf{r}d} \\ &+ \Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{G}\Phi_{\mathbf{rr}}^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{H}\right)^{\dagger}\left(\mathbf{f} - \mathbf{H}^{\mathrm{T}}\tilde{\delta}\right)\end{aligned} \tag{A4}$$

written in a compact form. Note that $\mathbf{f} \neq \mathbf{H}^{\mathrm{T}}\tilde{\delta}$. Otherwise, it would imply $\mathbf{w} = \mathbf{0}$ from the cost function in Eq. (9).

APPENDIX B: ADAPTIVE SOLUTION DERIVATION

The adaptive solution of $\mathbf{w}$ in the proposed method can be found using the LMS method as

$$\begin{aligned}\mathbf{w}(n+1) &= \mathbf{w}(n) - \mu\nabla_{\mathbf{w}}J \\ &= \mathbf{w}(n) - \mu\mathbf{G}^{\mathrm{T}}\left(\Phi_{\mathbf{xx}}\mathbf{u} + \mathbf{H}\lambda\right).\end{aligned} \tag{B1}$$

Using Eq. (3i), the constraint has

$$\begin{aligned}\mathbf{f} &= \mathbf{H}^{\mathrm{T}}\mathbf{u}(n+1) \\ &= \mathbf{H}^{\mathrm{T}}[\mathbf{G}\mathbf{w}(n+1) + \tilde{\delta}] \\ &= \mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{w}(n) - \mu\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\left[\Phi_{\mathbf{xx}}\mathbf{u} + \mathbf{H}\lambda(n)\right] + \mathbf{H}^{\mathrm{T}}\tilde{\delta}.\end{aligned} \tag{B2}$$

Putting $\lambda(n)$ from Eq. (B2) in Eq. (B1), it becomes

$$\begin{aligned}\mathbf{w}(n+1) = \mathbf{w}(n) &- \mu\mathbf{G}^{\mathrm{T}}\Phi_{\mathbf{xx}}\mathbf{u} \\ &- \mathbf{G}^{\mathrm{T}}\mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{H})^{\dagger}\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{w}(n) \\ &+ \mu\mathbf{G}^{\mathrm{T}}\mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{H})^{\dagger}\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\Phi_{\mathbf{xx}}\mathbf{u} \\ &+ \mathbf{G}^{\mathrm{T}}\mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{H})^{\dagger}(\mathbf{f} - \mathbf{H}^{\mathrm{T}}\tilde{\delta}). \end{aligned} \tag{B3}$$

511 Using Eq. (2) and rearranging Eq. (B3), it becomes

$$\mathbf{w}(0) = \mathbf{q}, \tag{B4a}$$

512

$$\mathbf{w}(n+1) = \mathbf{P}\left[\mathbf{w}(n) - \mu\mathbf{G}^{\mathrm{T}}\mathbf{x}(n)e(n)\right] + \mathbf{q}, \tag{B4b}$$

513 where

$$\mathbf{P} = \mathbf{I} - \mathbf{G}^{\mathrm{T}}\mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{H})^{\dagger}\mathbf{H}^{\mathrm{T}}\mathbf{G}, \tag{B5a}$$

514

$$\mathbf{q} = \mathbf{G}^{\mathrm{T}}\mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{H})^{\dagger}(\mathbf{f} - \mathbf{H}^{\mathrm{T}}\tilde{\delta}). \tag{B5b}$$

515

516 **Acknowledgment**

517 The authors wish to thank the anonymous reviewers for their helpful comments and
518 suggestions. The authors would also like to thank Jacob Donley and Sarmad Malik for their
519 assistance with this project.

520

Aboulnasr, T., and Mayyas, K. (**1997**). "A robust variable step-size lms-type algorithm: analysis and simulations," IEEE Transactions on Signal Processing **45**(3), 631–639, doi: mu10.1109/78.558478.

Acclivity (**2006**). "Henry5.mp3," Freesound muhttps://freesound.org/people/acclivity/sounds/24096/ (Last viewed March 18, 2021).

Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," The Journal of the Acoustical Society of America **58**(1), 214–222, muhttps://doi.org/10.1121/1.380648, doi: mu10.1121/1.380648.

Cartes, D. A., Ray, L. R., and Collier, R. D. (**2002**). "Experimental evaluation of leaky least-mean-square algorithms for active noise reduction in communication headsets," The Journal of the Acoustical Society of America **111**(4), 1758–1771.

Chang, C.-Y., Siswanto, A., Ho, C.-Y., Yeh, T.-K., Chen, Y.-R., and Kuo, S. M. (**2016**). "Listening in a noisy environment: Integration of active noise control in audio products," IEEE Consumer Electronics Magazine **5**(4), 34–43, doi: mu10.1109/MCE.2016.2590159.

Cheer, J., Patel, V., and Fontana, S. (**2019**). "The application of a multi-reference control strategy to noise cancelling headphones," The Journal of the Acoustical Society of America **145**(5), 3095–3103, muhttps://doi.org/10.1121/1.5109394, doi: mu10.1121/1.5109394.

Chen, J., Benesty, J., Huang, Y., and Doclo, S. (**2006**). "New insights into the noise reduction wiener filter," IEEE Transactions on Audio, Speech, and Language Processing **14**(4), 1218–1234, doi: mu10.1109/TSA.2005.860851.

Cox, H., Zeskind, R., and Owen, M. (**1987**). "Robust adaptive beamforming," IEEE Transactions on Acoustics, Speech, and Signal Processing **35**(10), 1365–1376, doi: mu10.1109/TASSP.1987.1165054.

Dalga, D., and Doclo, S. (**2011**). "Combined feedforward-feedback noise reduction schemes for open-fitting hearing aids," in *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 185–188, doi: mu10.1109/ASPAA.2011.6082318.

Doclo, S., Kellermann, W., Makino, S., and Nordholm, S. E. (**2015**). "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," IEEE Signal Processing Magazine **32**(2), 18–30, doi: mu10.1109/MSP.2014.2366780.

Donley, J., Tourbabin, V., Lee, J.-S., Broyles, M., Jiang, H., Shen, J., Pantic, M., Ithapu, V. K., and Mehra, R. (**2021**). "Easycom: An augmented reality dataset to support algorithms for easy communication in noisy environments," .

Elliott, S. J. (**2000**). *Signal processing for active control* (Academic Press), p. 511.

Elliott, S. J., Boucher, C. C., and Nelson, P. A. (**1992**). "The behavior of a multiple channel active control system," IEEE Transactions on signal processing **40**(5), 1041–1052.

Elliott, S. J., Jung, W., and Cheer, J. (**2018**). "Head tracking extends local active control of broadband sound to higher frequencies," Scientific reports **8**(1), 1–7.

Frost, O. L. (**1972**). "An Algorithm for Linearly Constrained Adaptive Array Processing," Proceedings of the IEEE **60**(8), 926–935, doi: mu10.1109/PROC.1972.8817.

Gannot, S., Vincent, E., Markovich-Golan, S., and Ozerov, A. (**2017**). "A consolidated perspective on multimicrophone speech enhancement and source separation," IEEE/ACM

Transactions on Audio, Speech, and Language Processing **25**(4), 692–730.

Hansen, C., Snyder, S., Qiu, X., Brooks, L., and Moreau, D. (**2012**). *Active Control of Noise and Vibration*, 2 ed. (CRC Press), p. 1554.

Hansen, P. C. (**2010**). *Discrete inverse problems: insight and algorithms* (SIAM).

Haykin, S. (**2002**). *Adaptive Filter Theory*, 4th ed. (Prentice Hall, Upper Saddle River, N.J.).

Kuo, S. M., and Morgan, D. R. (**1996**). *Active noise control systems : algorithms and DSP implementations* (Wiley), p. 389.

Lam, B., Shi, D., Gan, W.-S., Elliott, S. J., and Nishimura, M. (**2020**). "Active control of broadband sound through the open aperture of a full-sized domestic window," Scientific reports **10**(1), 1–7.

Li, J., Stoica, P., and Wang, Z. (**2003**). "On robust capon beamforming and diagonal loading," IEEE transactions on signal processing **51**(7), 1702–1715.

Liebich, S., Richter, J.-G., Fabry, J., Durand, C., Fels, J., and Jax, P. (**2018**). "Direction-of-Arrival Dependency of Active Noise Cancellation Headphones," Vol. ASME 2018 Noise Control and Acoustics Division Session presented at INTERNOISE 2018, muhttps://doi.org/10.1115/NCAD2018-6120, doi: mu10.1115/NCAD2018-6120, v001T08A003.

Mukai, R., Sawada, H., Araki, S., and Makino, S. (**2006**). "Frequency-domain blind source separation of many speech signals using near-field and far-field models," EURASIP Journal on Advances in Signal Processing **2006**, 1–13.

Patel, V., Cheer, J., and Fontana, S. (**2020**). "Design and Implementation of an Active Noise Control Headphone with Directional Hear-Through Capability," IEEE Transactions

on Consumer Electronics **66**(1), 32–40, doi: mu10.1109/TCE.2019.2956634.

Petersen, K. B., and Pedersen, M. S. (**2012**). "The matrix cookbook" muhttp://www2.compute.dtu.dk/pubdb/pubs/3274-full.html, version 20121115.

Rafaely, B., and Jones, M. (**2002**). "Combined feedback–feedforward active noise-reducing headset—the effect of the acoustics on broadband performance," The Journal of the Acoustical Society of America **112**(3), 981–989, muhttps://doi.org/10.1121/1.1501090, doi: mu10.1121/1.1501090.

Serizel, R., Moonen, M., Wouters, J., and Jensen, S. H. (**2010**). "Integrated active noise control and noise reduction in hearing aids," IEEE Transactions on Audio, Speech, and Language Processing **18**(6), 1137–1146, doi: mu10.1109/TASL.2009.2030948.

Shahbazpanahi, S., Gershman, A. B., Luo, Z.-Q., and Wong, K. M. (**2003**). "Robust adaptive beamforming for general-rank signal models," IEEE Transactions on Signal Processing **51**(9), 2257–2269.

Shen, X., Shi, D., and Gan, W.-S. (**2021**). "A wireless reference active noise control headphone using coherence based selection technique," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7983–7987, doi: mu10.1109/ICASSP39728.2021.9414683.

Shi, D., Gan, W.-S., Lam, B., and Wen, S. (**2020**). "Feedforward selective fixed-filter active noise control: Algorithm and implementation," IEEE/ACM Transactions on Audio, Speech, and Language Processing **28**, 1479–1492, doi: mu10.1109/TASLP.2020.2989582.

Shi, D., Lam, B., Ooi, K., Shen, X., and Gan, W.-S. (**2022**). "Selective fixed-filter active noise control based on convolutional neural network," Signal Processing **190**, 108317,

muhttps://www.sciencedirect.com/science/article/pii/S0165168421003546, doi: muhttps://doi.org/10.1016/j.sigpro.2021.108317.

Skogestad, S., and Postlethwaite, I. (**2005**). *Multivariable Feedback Control: Analysis and Design* (John Wiley & Sons, Inc., Hoboken, NJ, USA).

Tikhonov, A. N., and Arsenin, V. Y. (**1977**). "Solutions of ill-posed problems," New York **1**, 30.

Tobias, O. J., and Seara, R. (**2004**). "Leaky delayed lms algorithm: stochastic analysis for gaussian data and delay modeling error," IEEE transactions on signal processing **52**(6), 1596–1606.

Van Trees, H. L. (**2002**). *Optimum Array Processing* (John Wiley & Sons, Inc., New York, USA).

Van Veen, B., and Buckley, K. (**1988**). "Beamforming: a versatile approach to spatial filtering," IEEE ASSP Magazine **5**(2), 4–24, doi: mu10.1109/53.665.

Varga, A., and Steeneken, H. J. (**1993**). "Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems," Speech Communication **12**(3), 247–251, doi: muhttps://doi.org/10.1016/0167-6393(93)90095-3.

Vorobyov, S. A. (**2013**). "Principles of minimum variance robust adaptive beamforming design," Signal Processing **93**(12), 3264–3277.

Vorobyov, S. A., Gershman, A. B., and Luo, Z.-Q. (**2003**). "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," IEEE transactions on signal processing **51**(2), 313–324.

Wang, S., Tao, J., and Qiu, X. (**2017**). "Controlling sound radiation through an opening with secondary loudspeakers along its boundaries," Scientific reports **7**(1), 1–6.

Xiao, T., Qiu, X., and Halkon, B. (**2020**). "Ultra-broadband local active noise control with remote acoustic sensing," Scientific reports **10**(1), 1–12, doi: mu10.1038/s41598-020-77614-w.

Xu, B., and Miller, T. (**2019**). "Selective active noise cancellation based on directional reference signals," The Journal of the Acoustical Society of America **146**(4), 2794–2794, muhttps://doi.org/10.1121/1.5136678, doi: mu10.1121/1.5136678.

Zhang, H., and Wang, D. (**2021**). "Deep anc: A deep learning approach to active noise control," Neural Networks **141**, 1–10.

Zhang, L., and Qiu, X. (**2014**). "Causality study on a feedforward active noise control headset with different noise coming directions in free field," Applied Acoustics **80**, 36–44, muhttps://www.sciencedirect.com/science/article/pii/S0003682X1400005X, doi: muhttps://doi.org/10.1016/j.apacoust.2014.01.004.